

Fig. 1

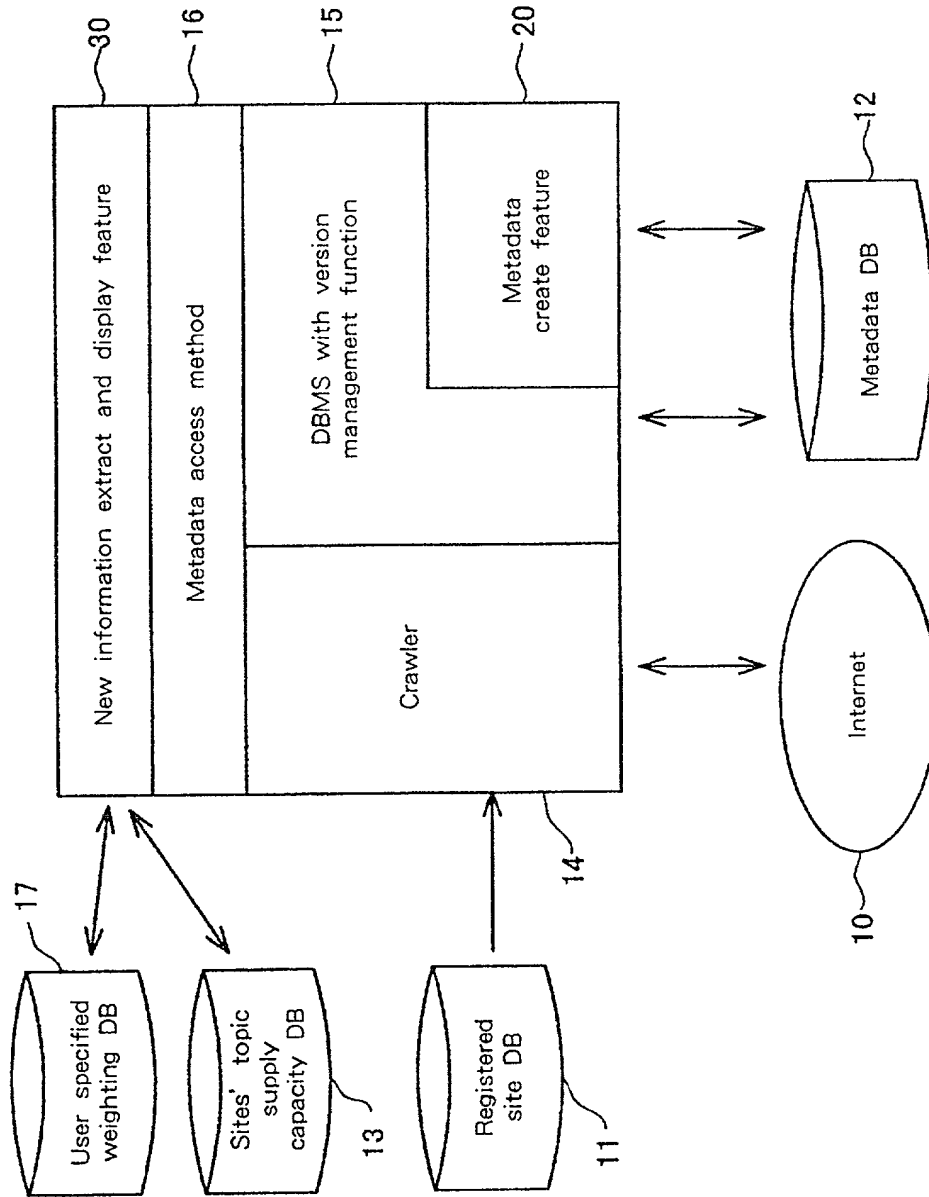


Fig. 2

3 / 11  
JP9-2DDD-DLBD (8728-512)

Example of registered sites

```
<site>
  <url>http://www.ibm.co.jp/</url>
  <name>IBM Japan Homepage</name>
</site>
<site>
  <url>http://www.ibm.com/</url>
  <name>IBM Corporation</name>
</site>
<site>
  <url>http://www.jp.ibm.com/shop/</url>
  <name>Japan IBM [shopping] </name>
</site>
<site>
  <url>http://www.jp.ibm.com/developerworks/</url>
  <name>developerWorks</name>
</site>
```

Fig. 3

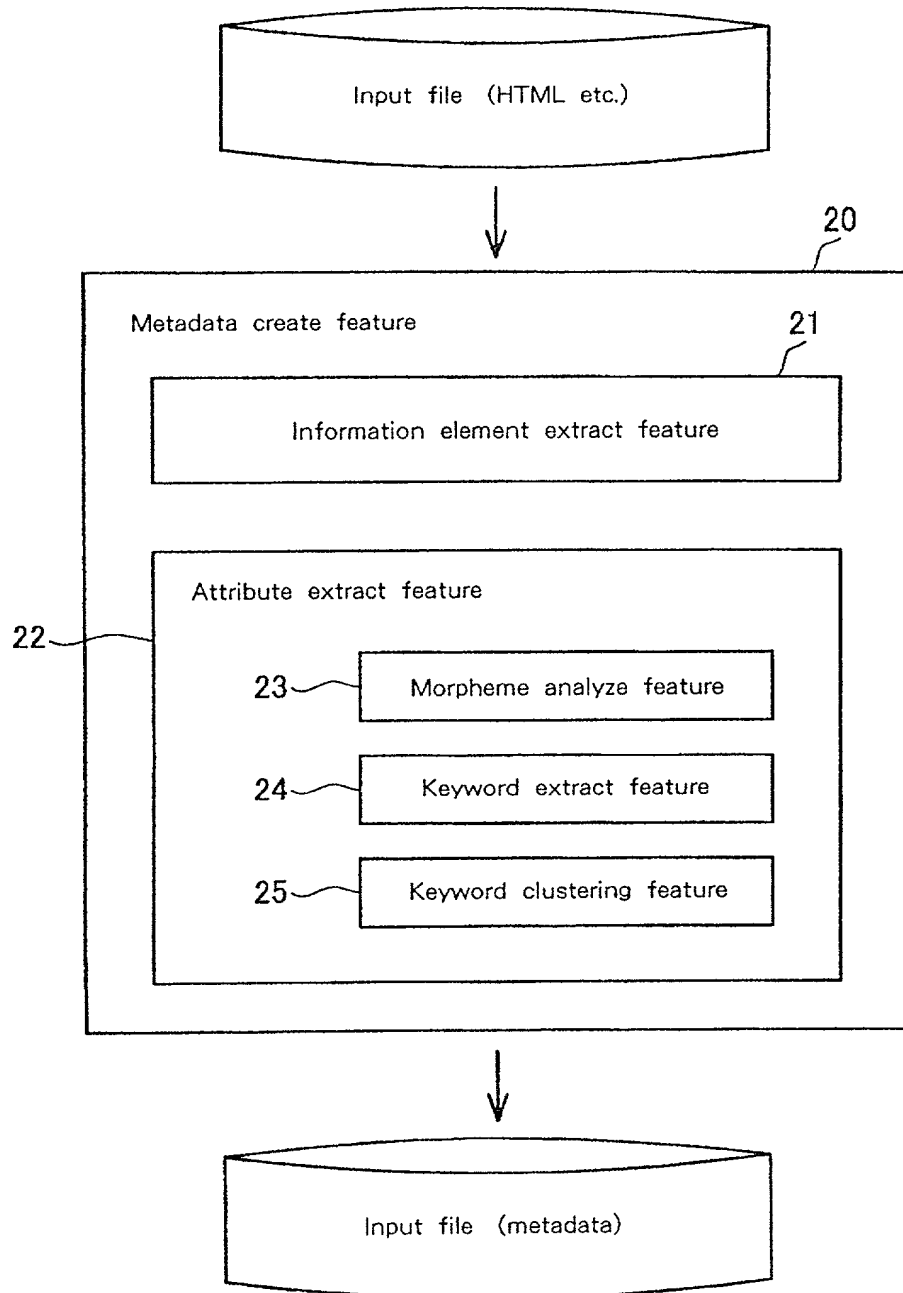


Fig. 4

### Example of Link

#### Expression in HTML file

```
<a href="http://www.ibm.com/jp/software/data/udb/v7/seminar.html">DB2 UDB V7 free seminar open </a>
```

#### Extracted information elements

```
<anchor>
// title of link
<DCtitle>DB2 UDB V7 free seminar open</DCtitle>
// URL
<url>http://www.ibm.com/jp/software/data/udb/v7/seminar.html</url>
// extracted keywords
<kwds>
//      keywords      //      clustering of keywords
<kwd><word>DB</word><class>T0</class></kwd>
<kwd><word>UDB</word><class>T0</class></kwd>
<kwd><word>V</word><class>T0</class></kwd>
<kwd><word> seminar </word><class>I3</class></kwd>
<kwd><word> open </word><class>I3</class></kwd>
<kwd><word> free </word><class>I3</class></kwd>
</kwds>
</anchor>
```

Fig. 5

Example of text block

Expression in HTML file

industry talk / e - column for this week

Extracted information elements

```
<text>
// text part
<DCdescription> industry talk / e - column for this week
</DCdescription>
<kwds>
<kwd><word>e</word><class>T0</class></kwd>
<kwd><word>industry</word><class>T0</class></kwd>
<kwd><word>column</word><class>11</class></kwd>
<kwd><word>talk</word><class>13</class></kwd>
<kwd><word>this week</word><class>11</class></kwd>
</kwds>
</text>
```

Fig. 6

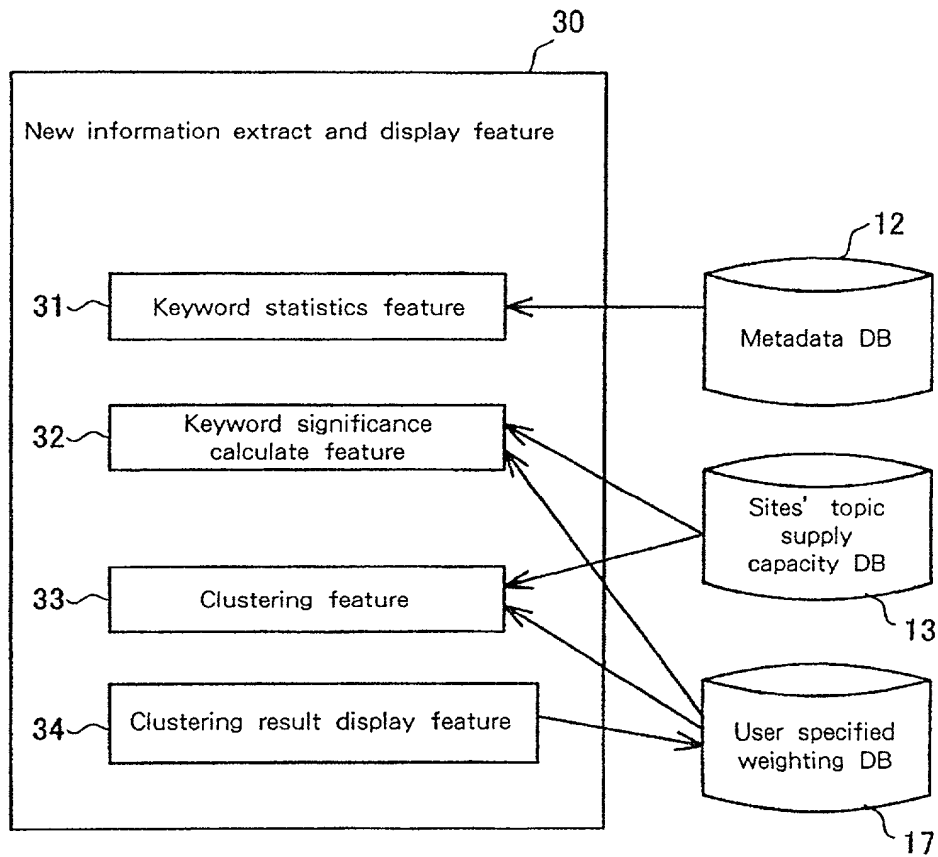


Fig. 7

● ● ●

8 / 11  
JP9-2000-D13D (8728-512)

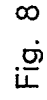


Fig. 8



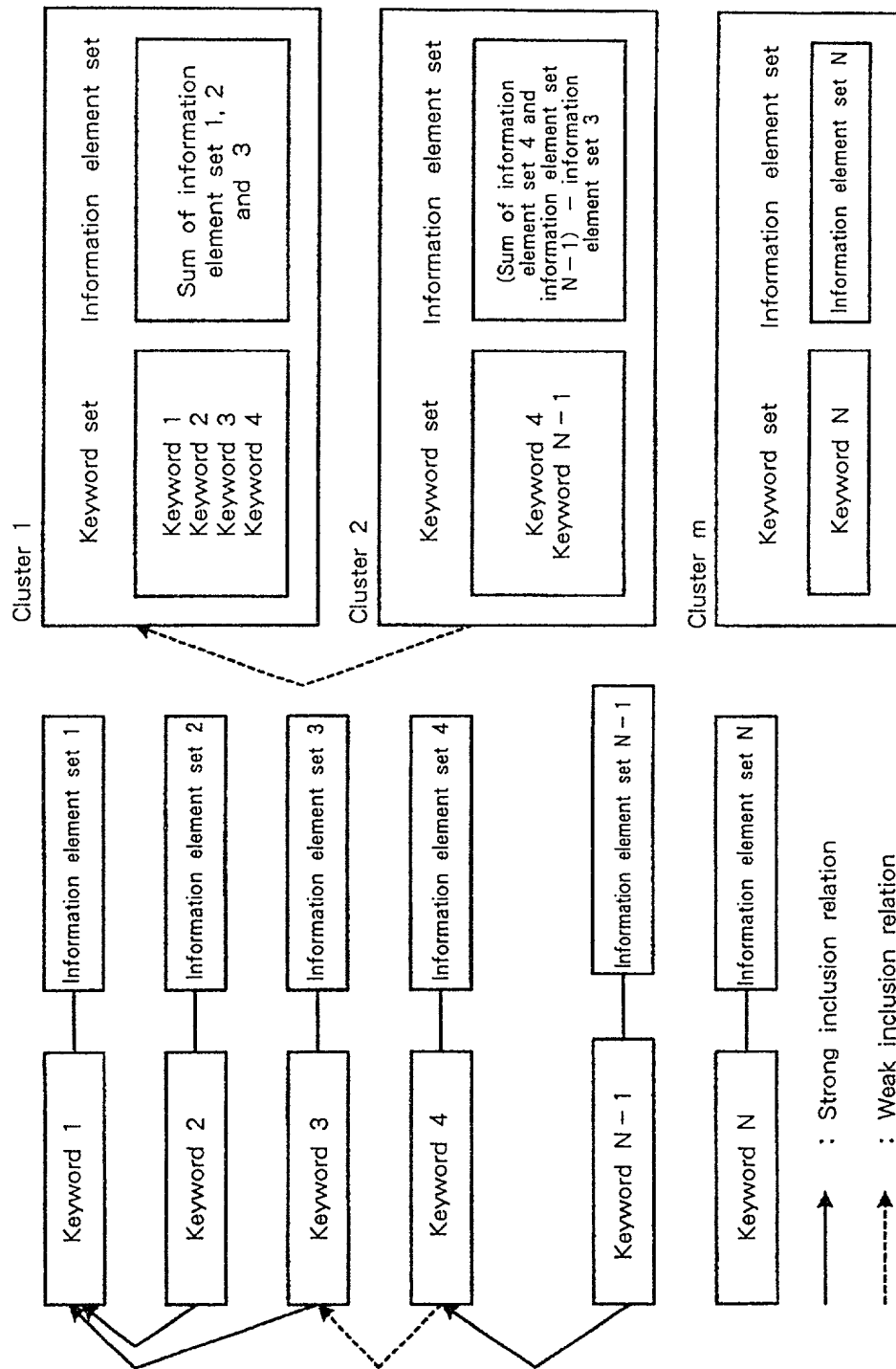
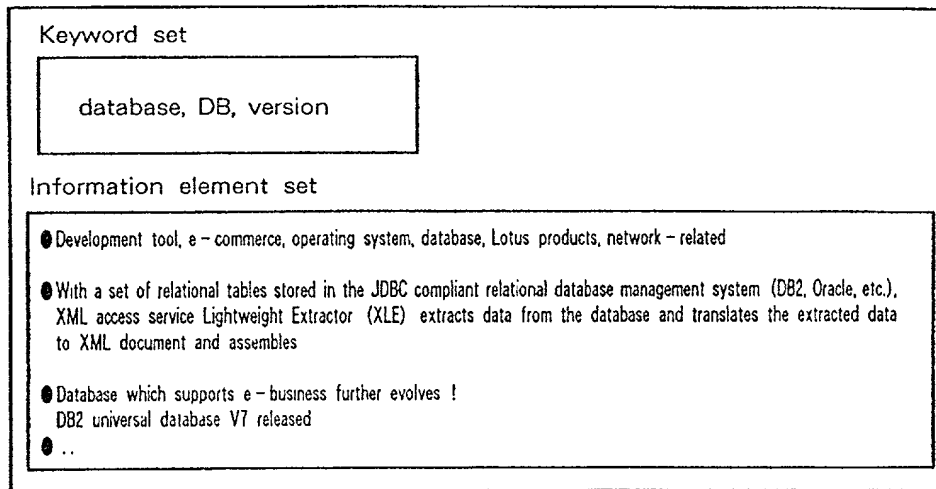
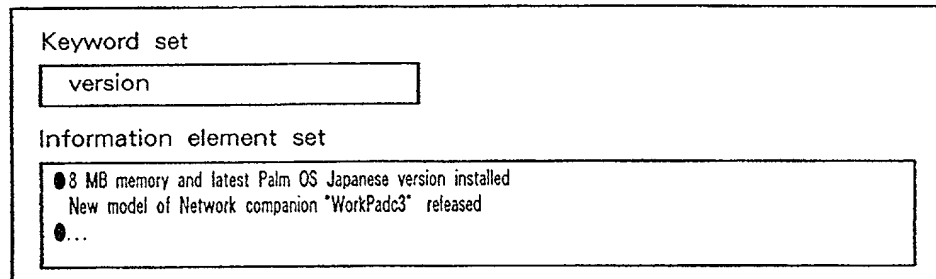


Fig. 9

### Cluster 1



### Cluster 2



### Cluster 3

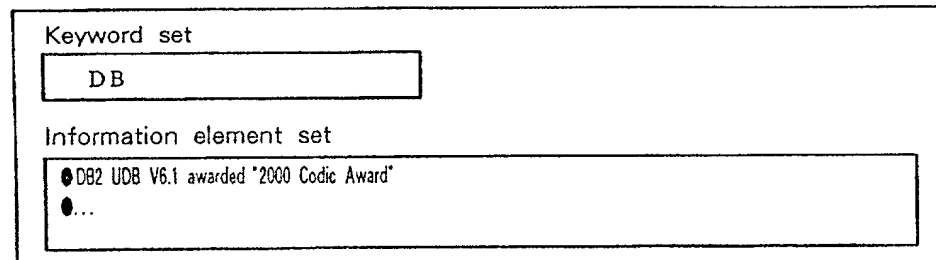


Fig. 10

